DEEP LEARNING WORKSHEET – 3

1.ANS) B) As number of hidden layers increase, model capacity increases

2.ANS) C) It normalizes (changes) all the input before sending it to the next layer

3.ANS) A) Network will not converge

4.ANS) D) All of these

5.ANS) C) (-4, -4, 3)

6.ANS) B) Simulate the network on a test dataset after every epoch of training. Stop training when the generalization error starts to increase

7.ANS) B) Stochastic Gradient Descent

8.ANS) A) Freeze all the layers except the last, re-train the last layer

9.ANS) B) TRAINING IS TOO SLOW, C) Restrict activations to become too high or low

10.ANS) B) SIGMOID

11.ANS) Activation functions are really important for an artificial neural network to learn and make sense of something really complicated and Non-linear complex functional mappings between the inputs and response variable.They introduce non-linear properties. Their main purpose is to convert an input signal of a node in a A-NN to an output signal, if we do not apply an activation function then the output signal would simply be a simple linear function

12.ANS) The input data is fed in the forward direction through the network. Each hidden layer accepts the input data, processes it as per the activation function and passes to the successive layer called forward propagation,

Back-propagation is an algorithm that computes the chain rule, with a speciﬁc order of operations that is highly eﬃcient. so, the total cost of backpropagation is roughly the same as making just two forward passes through the network. Compare that to the million and one forward passes of the previous method.

13.ANS) Gradient descent is a first-order iterative optimization algorithm for finding a local minimum of a differentiable function. To find a local minimum of a function using gradient descent,

Stochastic gradient descent refers to calculating the derivative from each training data instance and calculating the update immediately

Batch gradient descent is a variation of the gradient descent algorithm that calculates the error for each example in the training dataset, but only updates the model after all training examples have been evaluated. One cycle through the entire training dataset is called a training epoch

Mini-batch gradient descent is a variation of the gradient descent algorithm that splits the training dataset into small batches that are used to calculate model error and update model coefficients. Implementations may choose to sum the gradient over the mini-batch which further reduces the variance of the gradient.

14.ANS) Advantages of Mini-Batch Gradient Descent are Stable Convergence: Another advantage is the more stable converge towards the global minimum since we calculate an average gradient over n samples that results in less noise.

15.ANS) Transfer learning is the reuse of a pre-trained model on a new problem. It's currently very popular in deep learning because it can train deep neural networks with comparatively little data